Computing Interval Estimates for Components of Statistical Information with Respect to Judgements on Probability Density Functions

Victor G. Krymsky

Ufa State Aviation Technical University, Russia;

E-mail: kvg@mail.rb.ru

Kgs. Lyngby, Denmark, 2004

Imprecise Prevision Theory (IPT)

Starting Points: Fundamental Publications

[1], [2]

[1] **Walley P**., *Statistical reasoning with imprecise probabilities,* Chapman and Hall, New York, (1991);

[2] **Kuznetsov V**., *Interval statistical models*, Radio and Sviaz, Moscow, (1991) (in Russian).

Traditional Problem Formulation in the Framework of IPT

Constraints:

$$\rho(x) \geq 0, \quad \int_0^T \rho(x)\,dx = 1, \qquad \text{and}$$

$$\underline{a_i} \leq \int_0^T f_i(x)\rho(x)\,dx \leq \overline{a_i}, \quad i = 1,2,\ldots,n. \quad (1)$$

It is necessary to find:

$$\inf_{\rho(x)} \int_0^T g(x)\rho(x)\,dx \qquad \text{as well as}$$

$$\sup_{\rho(x)} \int_0^T g(x)\rho(x)\,dx \qquad (2)$$

subject to constraints (1).

Dual for the Initial Problem Statement

Let us find:

$$\underline{M}(g) = \sup_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^{n} (c_i \underline{a}_i - d_i \overline{a}_i) \right) \quad (3)$$

subject to $c_0 \in R$, $c_i, d_i \in R_+$ and for any $x \geq 0$, $i = 1, 2, \ldots, n$:

$$c_0 + \sum_{i=1}^{n} (c_i - d_i) f_i(x) \leq g(x). \quad (4)$$

And

$$\overline{M}(g) = \inf_{c_0, c_i, d_i} \left(c_0 + \sum_{i=1}^{n} (c_i \overline{a}_i - d_i \underline{a}_i) \right) \quad (5)$$

subject to $c_0 \in R$, $c_i, d_i \in R_+$ and for any $x \geq 0$, $i = 1, 2, \ldots, n$:

$$c_0 + \sum_{i=1}^{n} (c_i - d_i) f_i(x) \geq g(x). \quad (6)$$

Important Conclusion Concerning Optimal Solutions

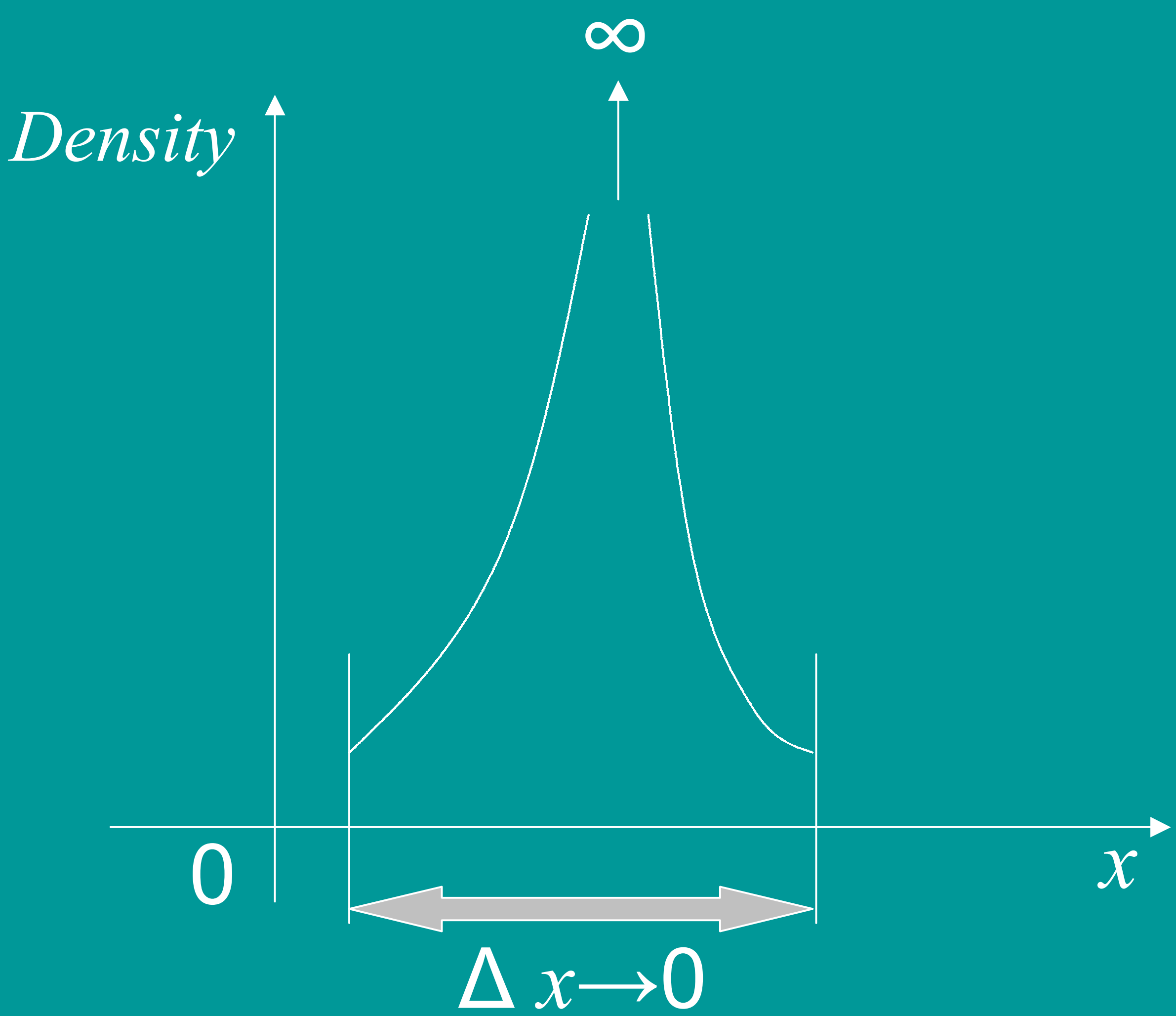
(L. Utkin and I. Kozine, [3])

[3] **Utkin L. and Kozine I.** *Different faces of the natural extension.* In: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications, ISIPTA '01, 2001, pp. 316-323.

Optimal solutions belong to a family of **DEGENERATE** distributions (such **probability densities are composed of** ***δ*****-functions**)

Distribution of Probabilistic Masses

Masses are concentrated in the fixed points:

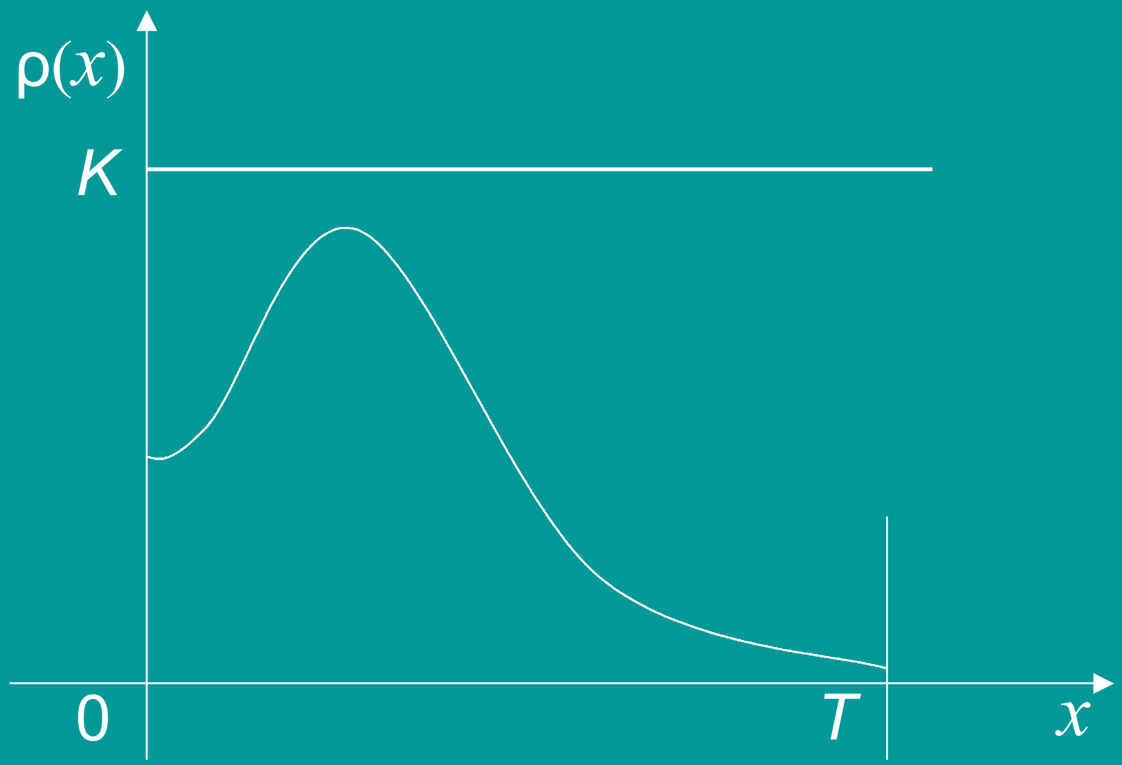


Use of Additional Judgements

Additional judgement can be reflected by inequality:

$\rho(x) \leq K = const,$ (7)

where $K \in R_+$ is such that $KT \geq 1.$



Main Goal (Theorem)

If there is no any finite interval $x \in [\alpha, \beta] \subseteq [0, T],\ \alpha < \beta,$ *for which function* $g(x)$ *can be represented in the form*

$$g(x) = h_0 + \sum_{i=1}^{n} h_i f_i(x), \qquad (8)$$

where $h_0, h_1, \ldots, h_n \in R,$ *then function* $\rho(x)$ *providing solution of optimization problem mentioned above, belongs to class of step-functions with minimum value equal to 0 and maximum value equal to K.*

Some Comments

To provide (8) the system

$$\left.\begin{aligned}
&h_0 + \sum_{i=1}^{n} h_i f_i(x) = g(x),\\
&\sum_{i=1}^{n} h_i f_i'(x) = g'(x),\\
&\sum_{i=1}^{n} h_i f_i''(x) = g''(x),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\sum_{i=1}^{n} h_i f_i^{(n)}(x) = g^{(n)}(x)
\end{aligned}\right\}$$

must have at least one solution which is independent on x in some interval $x \in [\alpha, \beta]$.

Applying Methodology of the Calculus of Variations

The inequalities

$$0 \le \rho(x) \le K \qquad (9)$$

should be excluded from direct consideration in order to allow operating in the open domain with the values of the function.

The requirement $\rho(x) \ge 0$ can be replaced by denoting

$$\rho(x) = z^2(x). \qquad (10)$$

The requirement $\rho(x) \le K$ can be reflected by equality

$$z^2(x) + v^2(x) = K, \qquad (11)$$

where $v(x)$ is newly introduced function.

Modified Formulation of the Problem

We would like to estimate

$$\inf_{z(x)} \int_0^T g(x) z^2(x)\,dx$$

and

$$\sup_{z(x)} \int_0^T g(x) z^2(x)\,dx \qquad (12)$$

subject to

$$z^2(x) + v^2(x) = K, \qquad (13)$$

$$\int_0^T z^2(x)\,dx = 1, \qquad (14)$$

$$\int_0^T f_i(x) z^2(x)\,dx \le \overline{a_i}, \qquad (15)$$

$$-\int_0^T f_i(x) z^2(x)\,dx \le -\underline{a_i}, \quad i = 1, 2, \ldots, n. \qquad (16)$$

Lagrange Approach

$$F(z,v)=g(x)z^2(x)$$

$$\downarrow$$

$$F^*(z,v)=g(x)z^2(x)+\lambda^*(x)\cdot\left(z^2(x)+v^2(x)\right)+$$

$$\lambda_0 z^2(x)+\sum_{i=1}^{n}\lambda_i f_i(x)z^2(x)-\sum_{i=n+1}^{2n}\lambda_i f_i(x)z^2(x)$$

Equations of Euler – Lagrange:

$$\frac{\partial F^*(z,v)}{\partial z}-\frac{d}{dx}\left(\frac{\partial F^*(z,v)}{\partial \dot{z}}\right)=0;$$

$$\frac{\partial F^*(z,v)}{\partial v}-\frac{d}{dx}\left(\frac{\partial F^*(z,v)}{\partial \dot{v}}\right)=0. \qquad (17)$$

The Necessary Conditions of Optimality

The equations look here as follows:

$$z(x)\cdot\left(g(x)+\lambda^*(x)+\lambda_0+\sum_{i=1}^{n}(\lambda_i-\lambda_{i+n})f_i(x)\right)=0;$$

(18)

$$\lambda^*(x)v(x)=0.$$

Let us fix any interval

$x\in[\alpha,\beta]\subseteq[0,T],\quad \alpha<\beta.$

Case 1. $z(x)\equiv 0$

inside the interval.

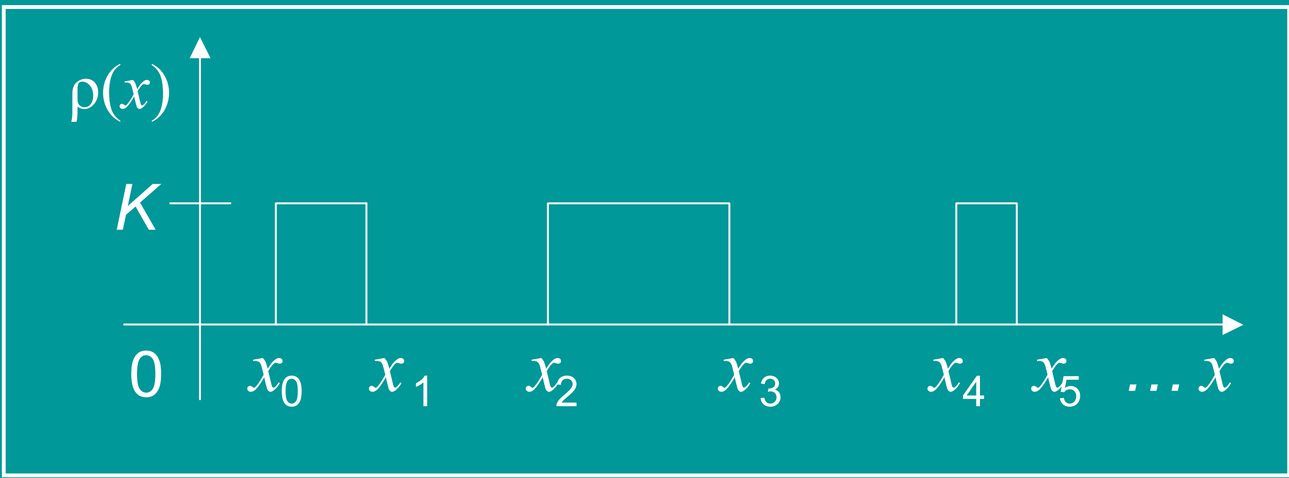
Then $v(x)=\sqrt{K}$ and $\lambda^*(x)=0.$

Case 2. $z(x)\neq 0,$ so

$v(x)=0$ and $z(x)=\sqrt{K}.$

Practical Implementation

Optimal probability density:



Denote:

$$G(x_j, x_{j+1}) = \int_{x_j}^{x_{j+1}} g(x)dx, \qquad (19)$$

$$\Phi_i(x_j, x_{j+1}) = \int_{x_j}^{x_{j+1}} f_i(x)dx, \quad i = 1,2,\ldots,n. \qquad (20)$$

Reformulation of the Problem Statement

We would like to estimate

$$\min_{x_0, x_1, \ldots} \left\{ K \cdot \sum_{j=0}^{m} G(x_{2j}, x_{2j+1}) \right\}, \qquad (21)$$

$$\max_{x_0, x_1, \ldots} \left\{ K \cdot \sum_{j=0}^{m} G(x_{2j}, x_{2j+1}) \right\} \qquad (22)$$

subject to

$$K \cdot \sum_{j=0}^{m} (x_{2j+1} - x_{2j}) = 1; \qquad (23)$$

$$\underline{a_i} \le K \cdot \sum_{j=0}^{m} \Phi_i(x_{2j}, x_{2j+1}) \le \overline{a_i}, \quad i = 1, 2, \ldots, n. \qquad (24)$$

Example 1

The information concerning a continuous random variable X is $\rho(x) \leq \Psi(x) = K \cdot I_{[0,T]}(x) < \infty$, where K, T are fixed positive numbers. What are **the bounds** for the **expectation $M(X)$**?

* * *

Let us choose m=0.

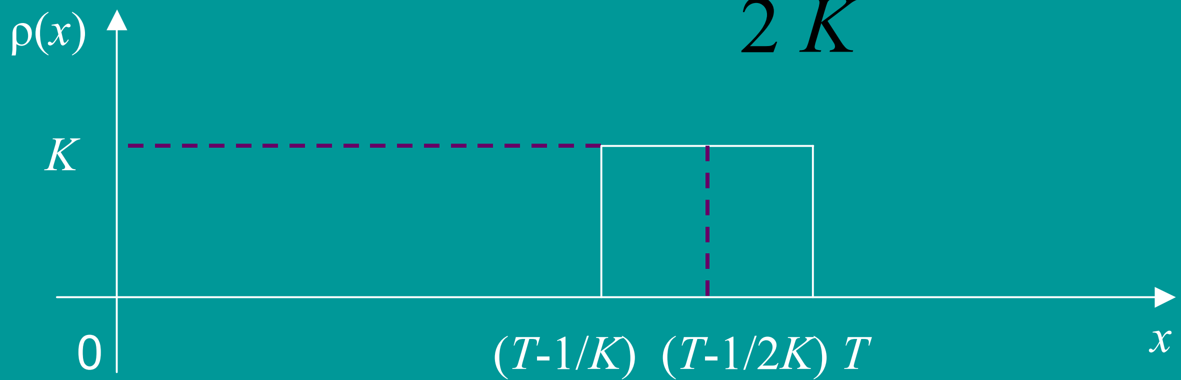
Objective function:

$$J = \int_0^T x\rho(x)dx = K \cdot \int_{x_0}^{x_1} xdx = K \cdot \frac{x_1^2 - x_0^2}{2}.$$

Solution of Optimization Problem

Lower and upper bounds of J interval:

$$\min \quad J = \frac{1}{2K}.$$



$$\max J = (T - 1/K) + \frac{1}{2K} = T - \frac{1}{2K}.$$

Example 2

We add the constraint:

$$\int_0^T f_1(x)\rho(x)\,dx = b,$$

where $f_1(x) = I_{[\underline{a},\bar{a}]}(x)$ is the indicator function.

Here also any finite interval of x values for which

$$g(x) = x = c_0 + c_1 I_{[\underline{a},\bar{a}]}(x)$$

cannot be found, so the theorem can be applied.

Further analysis shows, that $m=1$ is the best choice for such situation.

Example 2
(Continuation 1)

To provide $J \to \min$

we have to set:

(i) if $\underline{a} \geq \dfrac{1-b}{K}$:

$$\rho(x) = \begin{cases} K & \text{for } 0 \leq x \leq \dfrac{1-b}{K}; \\ 0 & \text{for } \dfrac{1-b}{K} < x < \underline{a}; \\ K & \text{for } \underline{a} \leq x \leq \underline{a} + \dfrac{b}{K}; \\ 0 & \text{for } \underline{\mathrm{a}} + \dfrac{b}{K} < x \leq T; \end{cases}$$

Example 2
(Continuation 2)

(ii) if $\underline{a} < \dfrac{1-b}{K}$:

$$\rho(x) = \begin{cases} K \text{ for } 0 \le x \le \underline{a} + \dfrac{b}{K}; \\ 0 \text{ for } \underline{a} + \dfrac{b}{K} < x < \bar{a}; \\ K \text{ for } \bar{\mathrm{a}} \le x \le \bar{a} + \dfrac{1 - \underline{a}K - b}{K}; \\ 0 \text{ for } \bar{a} + \dfrac{1 - \underline{a}K - b}{K} < x \le T. \end{cases}$$

As the result

$$\min J = \frac{1}{2K} + \frac{b\left(\underline{a}K - (1-b)\right)}{K} \quad \text{or}$$

$$\min J = \frac{1}{2K} + \frac{\bar{a}K - (\underline{a}K + b)\cdot\left(1 + (\bar{a} - \underline{a})K - b\right)}{K}.$$

Acknowledgements

- The research was initiated by Dr. Igor Kozine of Risø National Laboratory, Denmark, whose kind attention to this work is gratefully acknowledged.
- The work was partially supported by the grant T02-3.2-346 of Russian Ministry for Education which is also acknowledged.